

National Longitudinal Study of Adolescent Health

Wave III Education Data Linking Component

Chandra Muller, Jennifer Pearson, Catherine Riegle-Crumb,
Jennifer Harris Requejo, Kenneth A. Frank, Kathryn S. Schiller,
R. Kelly Raley, Amy G. Langenkamp, Sarah Crissey,
Anna Strassmann Mueller, Rebecca Callahan,
Lindsey Wilkinson, and Samuel Field



Carolina Population Center
University of North Carolina at Chapel Hill

August 2007

This research was funded by a grant from the National Institute of Child Health and Human Development under grant R01 HD40428-02 to the Population Research Center, University of Texas at Austin; Chandra Muller (PI) and the National Science Foundation grant number REC-0126167, Chandra Muller (PI). Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524 (addhealth@unc.edu).

WAVE III EDUCATION DATA LINKING COMPONENT

I. INTRODUCTION

A. Component Description

Complete documentation of the importance of the Linking variables in facilitating analyses using AHAA and Add Health data concurrently is available on the official AHAA and Add Health websites, <http://www.prc.utexas.edu/ahaa/> and <http://www.cpc.unc.edu/projects/addhealth>.

B. Summary of Linking Variables

The Linking indicators are designed to help analysts link student transcript data to academic or school course-taking years and to the Add Health surveys. Two types of Linking indicators were constructed, school-year and grade-level indicators. These indicators allow analysts to discern when each student was in high school, the duration of each student's high school career, what grade level each student was in during the first Add Health survey year, and each student's transcript-indicated grade level when high school course-taking began. Thus, these indicators are important for analysts interested in using the Academic Courses, Academic Networks, and Curriculum component indicators in conjunction with Add Health survey data.

C. Organization of Document

The following sections describe the school-year and grade-level Linking indicators. This description includes a discussion of their analytical importance, their interrelationship with constructed indicators included in other AHAA study components, and issues analysts need to be aware of when using them in analyses. Also presented are the conventions used to name Linking indicators, the missing codes assigned to them, and additional information about the Linking data file, [edulink](#).

II. CONSTRUCTED LINKING VARIABLES

A. Section 1: School-Year Indicators

ELYEAR0 – ELYEAR6

These variables refer to the school years in which students were taking high school courses, and allow analysts to infer the duration of their high school careers. **ELYEAR0** refers to high school level course-taking in the school year preceding each student's first year of actual high school enrollment. **ELYEAR1** is the school year that corresponds to the first year of high school course-taking data for each student.

ELYEAR2 is the second school year of high school course-taking data for each student, and so forth. A small number of students had more than four years of high school course-taking data. **ELYEAR5** refers to the fifth year of course-taking data, and **ELYEAR6** refers to the sixth year of course-taking. Students without course-taking data in the fifth and/or sixth years were assigned missing values for **ELYEAR5** and **ELYEAR6**. Approximately forty students had more than six years of course-taking data. For this small group of students, their assigned valid value for **ELYEAR6** does not correspond to their last year of course-taking, but instead refers to all their years of course-taking from the sixth year forward.

It is important to note that while school years run from the autumn of one year to the spring of the following year, school year values correspond to the autumn of that school year. Therefore, a value of 1994 on the **ELYEAR1** variable corresponds to the school year 1994-95.

In general, **ELYEAR1** corresponds to students' 9th grade or freshman year (see the discussion of **ELY1NINE** below), **ELYEAR2** corresponds to their 10th grade or sophomore year and so forth. Yet careful examination of students' transcripts revealed that the grade level assigned to students' courses sometimes varies within a given year. For example, some courses in a student's second year of high school might be labeled as 9th grade on their transcript and others labeled as 10th grade (referring to the understood level of the course). The meaning of grade level in high school is further complicated by the fact that some students take courses for longer than four years, and others for less. Additionally, the meaning of grade retention in high school is not clear, as students are unlikely to be 'held back' and required to retake an entire year's worth of courses in the same manner as students in primary school. In order to avoid the complications these issues raise for determining the correspondence between student course-taking and grade level, a numeric scheme (i.e., the first year of course-taking, the second year, etc.) was utilized for the **ELYEAR0-6** variables to indicate students' course-taking years. Indicators are provided, however, to supply analysts with information about the grade level of students' courses recorded on their transcripts. (See the discussion of **ELY1NINE** and **ELGLV945** below.)

Because **ELYEAR1** represents the first year of actual high school enrollment for all students whose transcripts were collected (and, therefore, by definition generally have recorded information for at least their first year of course-taking), there are no missing values for this variable. For subsequent years (**ELYEAR2-6**) a missing value of 9992 indicates that the student has graduated from or is no longer in high school, or at the very least, that there is no additional course-taking information available for that student. For **ELYEAR0** (the year preceding high school enrollment), a missing value of 9992 is assigned to students who did not have year 0 information recorded on their transcripts.

A small number of students have one or two year-long gaps in their high school transcripts when no courses were recorded. These gaps could be the result of transcript error, or could indicate that students were not attending high school when the gaps occur on their transcripts. These students have valid values for the **ELYEAR** variables

corresponding to the timing of the gaps on their transcripts, but are assigned a missing value of 9992 for all constructed Academic Courses indicators (such as grade point average) during this time-frame. Consequently, the number of students with a missing value of 9992 on an **ELYEAR** variable in a given year is slightly smaller than the number of students with a missing value of 9992 on any corresponding constructed Academic Courses indicator. This difference indicates the number of students who have no course-taking information in that year, but who do have course-taking information recorded in one or more subsequent years.

For clarification purposes, it is important to note that all students assigned a missing value of 9992 for **ELYEAR0** are also assigned this missing value for all year **0** Academic Courses indicators.

All of the constructed Academic Courses indicators have names ending with a number as the last character, which directly correspond to the numeric schema used for the school-year indicators. For example, the variable **EAMGPA1** represents students' math GPA for their first year, **ELYEAR1**, of high school course-taking. (The Curriculum and Local Positions indicators were similarly constructed.)

ELMAT945

Due to the multi-cohort design of Add Health, students' high school careers overlap differently with the survey years. For example, students who entered Add Health as 11th or 12th graders, have academic information from the AHAA study that precedes their survey responses. For those students who entered Add Health as 7th and 8th graders, most of their survey data occurs before they entered high school, and therefore, precedes all of the AHAA constructed Academic Course indicators created from their high school transcripts. The **ELMAT945** variable allows analysts to easily match students' high school course-taking information to the school year 1994-95, when Add Health's In-School Survey was conducted, thus providing a link between Add Health survey data and AHAA transcript data. Values from 0-6 are applied to this variable to indicate which school year corresponds to 1994-95 for that student. For example, a value of 3 would indicate that the student was in their third year of course-taking, their junior year, during the school year of 1994-95. Furthermore, if an analyst is interested in examining students' overall grade point average at the same time point as a student's response to a survey question from the In-School Survey, the value under **ELMAT945** would enable the analyst to determine which of the students' overall grade point average variables to use (**EAOGPA1**, **EAOGPA2**, **EAOGPA3**, **EAOGPA4**, **EAOGPA5**, or **EAOGPA6**).

Additionally, because all of the questions answered in Wave I refer to students' experiences during the school year 1994-95, the **ELMAT945** variable also indicates the school year that these students referred to when responding to Wave 1 questions. Analysts may still wish to exercise caution with those Add Health students who responded to the Wave I survey during the end of the summer of the 1994-95 school year, as well as those who answered it in the beginning months of the 1995-96 school year.

A small number of students' last year of course-taking preceded 1994-95, and are assigned a missing code of 9994 on **ELMAT945**. Because Add Health began in that year, it is likely that these students were in high school in 1994-95, but have incomplete transcript information. Also, there are a few students whose value for **ELYEAR6** is 1993 (for the school year 1993-94), but their value on **ELMAT945** is 6 (indicating that the 6th year of course-taking occurred during 1994-95). Note that this is due to the convention of collapsing years 7-12 course-taking information with year 6 information, and that these students were actually in their seventh year of course-taking in 1994-95.

B. Section 2: Grade-Level Indicators

ELGLV945, and ELY1NINE

ELGLV945 indicates students' grade level in the school year 1994-95 (the year of the In-School Survey of Add Health). Because each course a student took in a given year is assigned a grade level on their transcript, this variable was calculated as the mean grade level of all courses taken in that year. For most students, all of the courses taken in a given school year had the same grade level, and therefore they have a whole number value for **ELGLV945**. However, some students had courses marked with different grade levels during the same school year, and therefore do not have a whole number value on **ELGLV945**.

It should be noted that this variable is calculated using only information from students' high school transcripts. Official school records of students' grade level may differ from students' survey reports of their grade level. Schools may treat grade level as a function of credits accrued toward graduation requirements, and therefore take account of course failures or difficulty of courses when assigning grade level. By contrast, students may consider grade level as a function of their length of time in school. In the majority of cases, there is direct correspondence between grade levels reported by students and those reported by schools, yet discrepancies do exist. For these reasons, analysts may choose to compare values on **ELGLV945** to students' self-reported grade level from the Add Health surveys.

While **ELGLV945** provides information about students' transcript-indicated grade level in the school year 1994-95, **ELY1NINE** is a dichotomous indicator of whether or not students' **ELYEAR1**, or the first year of high school course-taking data (of actual high school enrollment), generally corresponds to their 9th grade year. This is determined by the mean grade level assigned to all courses students took in **ELYEAR1**. For students whose average grade level was exactly nine (which was the large majority of students), or greater than nine but less than ten, we can generally conclude that the first year of course-taking data is their 9th-grade year. These students have a value of 1 on **ELY1NINE**.

Students with a value of 0 on **ELY1NINE** had a transcript-indicated average grade level that was ten or higher for their first year of course-taking data. This could

indicate that the student's transcript was incomplete and did not include their 9th grade courses, or instead that the student began high school at a relatively accelerated position, taking 10th grade courses or higher.

Analysts familiar with Add Health are aware that it includes several high schools with a grade range of only 10th-12th grades. Because high school transcripts nationwide are required to include students' course-taking information from grades 9-12, students attending these schools still have 9th grade courses listed on their transcripts. Thus, analysts should not speculate or conclude that students with a value of 0 on **ELY1NINE** are from high schools containing only 10th – 12th grades.

Because a mean grade level in 1994-95 could not be computed for students whose first year of high school course-taking information took place after 1994-95 (i.e., the 7th and 8th grade cohorts), these students are assigned a missing value of 9993. Likewise, a mean grade level could not be calculated for the small number of students whose last year of recorded course-taking occurred before the 1994-95 academic year. These students are assigned a missing value of 9994. See the section of missing codes for more information about specific missing values for Linking variables.

III. NAMING CONVENTIONS

Linking variables were labeled according to a standard format:

- A.** Each variable consists of up to but not exceeding 8 upper-case characters.
- B.** The initial character of all Linking variables is **E**, serving as a reference to the educational component of Add Health (AHAA).
- C.** The second character of all Linking variables, the letter **L**, indicates that these variables belong to the Linking component of the broader Add Health/AHAA data set.
- D.** The next characters in Linking variable names further specify the subject matter represented by each variable.
- E.** The last characters denote a concept of time. For example, for the year variables, the last character in these variables' names indicates a specific year of high school course-taking (i.e., **0**, **1**, **2**, **3**, **4**, **5**, or **6** corresponding to course-taking before and throughout the first through sixth year of course-taking, respectively).

The following representative examples serve as a guide for interpreting the nomenclature used to describe Linking variables:

1. Example of a school-level indicator:

ELYEAR1 – *the school year that corresponds to the first year of high school course taking*

This variable's initial two letters, **EL**, identify it as part of the education module (AHAA), and more specifically the Linking component, of Add Health. The next four characters, **YEAR**, all relate to the variable's subject, the school year in which students were taking high school courses. The last character, the number **1**, shows that this variable contains information about the school year that corresponds to each student's first year of high school course-taking.

2. Example of a grade-level indicator:

ELGLV945 – *student's grade level in the school year 1994-95.*

As in the first example, the initial two letters, **EL**, identify this variable as part of the Linking component within the educational unit, AHAA, of Add Health. The next three characters, **GLV**, describe this variable as pertaining to student's grade level. The last three characters, **945**, inform the analyst that this variable refers to the school year 1994-95.

IV. MISSING CODES

A. *Description of Missing Codes*

1. 9991: No transcript data exists for a given student.
2. 9992: Student has graduated from or is no longer in high school, or there is no additional course-taking information available for that student.
3. 9993: Respondent's first year of high school course-taking occurred after 1994-95.
4. 9994: Respondent's last year of course-taking preceded 1994-95 (the school year when the In-School Survey and Wave I were conducted).
5. 9995: No courses are recorded on a student's transcript in the year 1994-95.

B. *Variable-Specific Considerations*

1. Year Variables (ELYEAR0-ELYEAR6)

Students were assigned the missing value 9992 for **ELYEAR0** (variable that identifies the school years in which students were taking high school courses), if they do

not have any courses listed on their transcript that were taken prior to enrollment in high school. These students are also assigned missing values for all year **0** Academic Courses indicators.

Students assigned a missing value on year **1-6** variables are those whose transcript information ends in the specified year – generally indicating that the student is no longer traceable in high school. However, a small sub-set of students had a one- to two-year gap in their transcript files in which no courses were recorded. This occurrence could be due to a range of indeterminate possibilities including transcript errors or omissions, incorrect coding, or students' temporary withdrawal from high school. 'Valid' rather than 'missing' values were assigned for relevant year variables for these students, facilitating an analytical distinction between them and those students whose transcripts indicate a permanent separation from high school. Importantly, these students with holes in their transcript records were assigned missing values for all Academic Courses variables (such as grades), allowing analysts to separate them from the rest of the AHAA student population. In addition, assigning these students valid values for year variables keeps intact the temporal sequential feature and comparability of year variables (e.g., year **2** information referring to the second year of course-taking information for all AHAA students and succeeding year **1** in a one-year time interval across students).

2. Students Grade Level during 1994-95 (ELGLV945)

Because this variable conveys information about the mean grade level of courses taken during 1994-95, missing values are assigned to students whose high school course-taking began after or ended before this academic calendar year. Specifically, the missing value 9993 indicates that a student's year **1** information took place after 1994-95, thus a mean grade level in 1994-95 could not be computed for this student. Likewise, a missing value of 9994 signifies that a student's last year of recorded course-taking occurred before the 1994-95 school year, and, therefore, a mean grade level for 1994-95 could not be calculated for this student.

V. DATA FILE INVENTORY

The data file containing the school-year and grade-level indicators is [edulink](#). All students with transcript data are included in this data file. (See the student-level disposition section in the Primary component users' guide for information about which students have transcript data.)

VI. QUICK REFERENCE GUIDE

A. School-Year Indicators

ELYEAR(0–6) These variables refer to the school years in which students were taking high school courses, and allow analysts to infer the duration of their high school careers.

ELMAT945 The ELMAT945 variable allows analysts to easily match students' high school course-taking information to the school year 1994-95, when Add Health's In-School Survey was conducted, thus providing a link between Add Health survey data and AHAA transcript data.

B. Grade-Level Indicators

ELGLV945 Indicates students' grade level in the school year 1994-95 (the year of the In-School Survey of Add Health).

ELY1NINE A dichotomous indicator of whether or not students' ELYEAR1, or the first year of high school course-taking data (of actual high school enrollment), generally corresponds to their 9th grade year.

Education Data - Linking Component

Education Data - Linking Component

aid		Str 8	Respondent Identifier NOTE: Smallest 5 and largest 5 values are displayed.
Frequency	Percent	Value	Label
1	0.0%	10316654	
1	0.0%	10316952	
1	0.0%	10506342	
1	0.0%	10606128	
1	0.0%	11316754	
12227	99.9%	11316958-99719976	NOTE: Range of values omitted from display
1	0.0%	99886991	
1	0.0%	99886994	
1	0.0%	99886995	
1	0.0%	99886996	
1	0.0%	99886999	

elyear0		Num 2	School Year Preceding Student's First Year of High School
Frequency	Percent	Value	Label
5	0.0%	1989	1989
129	1.1%	1990	1990
141	1.2%	1991	1991
133	1.1%	1992	1992
124	1.0%	1993	1993
120	1.0%	1994	1994
114	0.9%	1995	1995
11471	93.7%	9992	no course taking data in year 0

elyear1		Num 2	1st Year of High School Course Taking
Frequency	Percent	Value	Label
9	0.1%	1987	1987
5	0.0%	1988	1988

12	0.1%	1989	1989
93	0.8%	1990	1990
1933	15.8%	1991	1991
2353	19.2%	1992	1992
2339	19.1%	1993	1993
2107	17.2%	1994	1994
1602	13.1%	1995	1995
1631	13.3%	1996	1996
103	0.8%	1997	1997
24	0.2%	1998	1998
26	0.2%	1999	1999

elyear2		Num 2	2nd Year of High School Course Taking
Frequency	Percent	Value	Label
8	0.1%	1988	1988
4	0.0%	1989	1989
11	0.1%	1990	1990
93	0.8%	1991	1991
1931	15.8%	1992	1992
2349	19.2%	1993	1993
2314	18.9%	1994	1994
2049	16.7%	1995	1995
1553	12.7%	1996	1996
1577	12.9%	1997	1997
84	0.7%	1998	1998
16	0.1%	1999	1999
248	2.0%	9992	no course taking data in year 2

elyear3		Num 2	3rd Year of High School Course Taking
Frequency	Percent	Value	Label
7	0.1%	1989	1989
4	0.0%	1990	1990
11	0.1%	1991	1991
92	0.8%	1992	1992

1925	15.7%	1993	1993
2307	18.9%	1994	1994
2212	18.1%	1995	1995
1956	16.0%	1996	1996
1470	12.0%	1997	1997
1518	12.4%	1998	1998
64	0.5%	1999	1999
671	5.5%	9992	no course taking data in year 3

elyear4		Num 2	4th Year of High School Course Taking
Frequency	Percent	Value	Label
4	0.0%	1990	1990
4	0.0%	1991	1991
10	0.1%	1992	1992
90	0.7%	1993	1993
1886	15.4%	1994	1994
2185	17.9%	1995	1995
2104	17.2%	1996	1996
1827	14.9%	1997	1997
1393	11.4%	1998	1998
1433	11.7%	1999	1999
25	0.2%	2000	2000
1276	10.4%	9992	no course taking data in year 4

elyear5		Num 2	5th Year of High School Course Taking
Frequency	Percent	Value	Label
4	0.0%	1992	1992
8	0.1%	1993	1993
69	0.6%	1994	1994
125	1.0%	1995	1995
124	1.0%	1996	1996
135	1.1%	1997	1997
94	0.8%	1998	1998
54	0.4%	1999	1999

11624	95.0%	9992	no course taking data in year 5
-------	-------	------	---------------------------------

elyear6		Num 2	6th Year of High School Course Taking
Frequency	Percent	Value	Label
3	0.0%	1993	1993
6	0.0%	1994	1994
17	0.1%	1995	1995
13	0.1%	1996	1996
27	0.2%	1997	1997
17	0.1%	1998	1998
21	0.2%	1999	1999
12133	99.2%	9992	no course taking data in year 6

elmat945		Num 2	Match Between ELYEAR1-6 and 1994-95
Frequency	Percent	Value	Label
2107	17.2%	1	year 1 = 1994-95
2314	18.9%	2	year 2 = 1994-95
2307	18.9%	3	year 3 = 1994-95
1886	15.4%	4	year 4 = 1994-95
69	0.6%	5	year 5 = 1994-95
9	0.1%	6	year 6 = 1994-95
3386	27.7%	9993	year 1 after 1994-95
159	1.3%	9994	last year of course taking data before 1995-95

elglv945		Num 8	Mean Grade Level of Courses in 1994-95 NOTE: Smallest 5 and largest 5 values are displayed.
Frequency	Percent	Value	Label
2086	17.0%	9	9th grade
69	0.6%	9.1	9.1
6	0.0%	9.2	9.2
5	0.0%	9.5	9.5
2	0.0%	9.6	9.6
4689	38.3%	9.7-11.8	NOTE: Range of values omitted from display
7	0.1%	11.9	11.9
1816	14.8%	12	12th grade

3386	27.7%	9993	year 1 after 1994-95
159	1.3%	9994	last year of course taking data before 1995-95
12	0.1%	9995	no courses recorded in 1994-95

ely1nine		Num 1	Grade Level Year 1
Frequency	Percent	Value	Label
172	1.4%	0	mean grade level of year 1 courses is greater than or equal to 10
12065	98.6%	1	mean grade level of year 1 courses is between 9 and 10